ASIGNATURA: ESTADÍSTICA Y PROCESOS ESTOCÁSTICOS
EXAMEN EXTRAORDINARIO 2015-16 **Duración:** 3 horas
FECHA: 12 de Julio de 2016
Fecha publicación notas: 15 de Julio de 2016
Fecha revisión examen: 18 de Julio de 2016
APELLIDOS: **NOMBRE:**
DNI: **TITULACIÓN:**

Ejercicio 1 *Un químico analiza muestras de agua del mar Mediterráneo para detectar la presencia de dos metales pesados: plomo y mercurio.*

Una muestra puede contener niveles altos de plomo (suceso A) o bajos y niveles altos de mercurio (suceso B) o bajos.

Se sabe que el 38 % de las muestras que se toman en el mar Mediterráneo tiene niveles altos de alguno de los dos metales, el 32 % tiene un nivel elevado de plomo y el 15 % tiene niveles elevados de mercurio.

(a) *(0.5 puntos) Si una muestra contiene un nivel alto de plomo ¿cuál es la probabilidad de que contenga un bajo nivel de mercurio?*

Solución.

Los datos que nos dan en el enunciado son

$$P(A \cup B) = 0.38, P(A) = 0.32, P(B) = 0.15$$

Nos piden:

$$P\left(\overline{B}/A\right) = \frac{P\left(\overline{B} \cap A\right)}{P\left(A\right)}$$

donde

$$P\left(\overline{B} \cap A\right) = P(A) - P\left(B \cap A\right) =$$

$$= P(A) - [P(A) + P(B) - P\left(B \cup A\right)] = 0.38 - 0.15 = 0.23.$$

Entonces,

$$P\left(\overline{B}/A\right) = \frac{0.23}{0.32} = 0.71875$$

(b) *(0.5 puntos) Si se toman dos muestras independientes, calcula la probabilidad de que en alguna de ellas se obtenga un nivel alto de plomo y un bajo nivel de mercurio.*

Solución.

Definimos la variable aleatoria X: número de muestras que contienen un nivel alto de plomo y bajo de mercurio, de las dos muestras tomadas.
Por definición, X sigue un modelo binomial de parámetros $n = 2$ y $p = P\left(A \cap \overline{B}\right) = 0.23$. Nos piden

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - 0.23)^2 = 0.4071$$

También podemos resolver este apartado así:

Definimos los sucesos M_i: la i-ésima muestra tomada en el Mediterráneo contiene un nivel alto de plomo y bajo nivel de mercurio, $i = 1, 2$. Sabemos que $P(M_1) = P(M_2) = P\left(A \cap \overline{B}\right) = 0.23$. Además, como las muestras son independientes, los sucesos M_1, M_2 son independientes. Entonces,

$$P(M_1 \cup M_2) = P(M_1) + P(M_2) - P(M_1) \cdot P(M_2) = 0.23 + 0.23 - 0.23^2 = 0.4071$$

(c) *(0.5 puntos) Se analizan muestras consecutivamente. Calcula la probabilidad de que la primera con niveles altos de los dos metales aparezca en tercer lugar.*

Solución.

Definimos la variable aleatoria

Y: número de muestras que hay que analizar para encontrar la primera con niveles altos de los dos metales.

Por definición, Y sigue un modelo geométrico de parámetro $p = P\left(A \cap B\right) = P(A) + P(B) - P(A \cup B) = 0.32 + 0.15 - 0.38 = 0.09$. Nos piden

$$P(Y = 3) = 0.09 \cdot 0.91^2 = 0.0745$$

Se dispone ahora de un conjunto de muestras procedentes de dos mares. El 40 % de ellas procede del mar Mediterráneo (suceso M) y el 60 % restante procede del mar Cantábrico (suceso C).

Para las muestras tomadas en el mar Cantábrico el 25 % tiene niveles altos de alguno de los dos metales, el 20 % tiene un nivel elevado de plomo y el 12 % tiene niveles elevados de mercurio.

(d) *(0.5 puntos) Si se elige una muestra al azar y resulta que contiene niveles altos de ambos tipos de metales ¿cuál es la probabilidad de que esta muestra haya sido tomada en el Mediterráneo?*

Solución.

Nos indican que $P(M) = 0.40$ y $P(C) = 0.60$ y que

$$P(A \cup B/C) = 0.25, P(A/C) = 0.30, P(B/C) = 0.12$$

Obsérvese además que las probabilidades que hemos utilizado en el apartado a) realmente son probabilidades condicionadas al suceso M, es decir,

$$P(A \cup B/M) = 0.38, P(A/M) = 0.32, P(B/M) = 0.15$$

Nos piden calcular

$$P(M/A \cap B) = \frac{P(A \cap B/M) \cdot P(M)}{P(A \cap B)} =$$

$$= \frac{P(A \cap B/M) \cdot P(M)}{P(A \cap B/M) \cdot P(M) + P(A \cap B/C) \cdot P(C)} =$$

$$= \frac{0.09 \cdot 0.40}{0.09 \cdot 0.40 + 0.07 \cdot 0.60} = 0.4615$$

En este cálculo hemos empleado el teorema de Bayes.

Ejercicio 2 *Una urna contiene 7 bolas rojas y 3 azules. Se seleccionan al mismo tiempo, de forma aleatoria, 5 bolas.*

(a) *(0.5 puntos) Determina la función de probabilidad de la variable aleatoria N que mide "el número de de bolas rojas obtenidas".*

Solución.

N puede tomar los valores $2, 3, 4$ y 5.

$$P(N=2) = \frac{\binom{7}{2}}{\binom{10}{5}} = \frac{1}{12}$$

$$P(N=3) = \frac{\binom{7}{3}\binom{3}{2}}{\binom{10}{5}} = \frac{5}{12}$$

$$P(N=4) = \frac{\binom{7}{4}\cdot 3}{\binom{10}{5}} = \frac{5}{12}$$

$$P(N=5) = \frac{\binom{7}{5}}{\binom{10}{5}} = \frac{1}{12}$$

(b) *(0.3 puntos) Calcula la función de distribución de N.*

Solución.

$$F(x) = P(N \leq x) = \begin{cases} 0 & \text{si} \quad x < 2 \\ P(N=2) = \frac{1}{12} & \text{si} \quad 2 \leq x < 3 \\ P(N=2) + P(N=3) = \frac{1}{2} & \text{si} \quad 3 \leq x < 4 \\ P(N=2) + P(N=3) + P(N=4) = \frac{11}{12} & \text{si} \quad 4 \leq x < 5 \\ 1 & \text{si} \quad 5 \leq x \end{cases}$$

(c) *(0.2 puntos) Calcula la media de N.*

Solución.

$$E[N] = 2\cdot\frac{1}{12} + 3\cdot\frac{5}{12} + 4\cdot\frac{5}{12} + 5\cdot\frac{1}{12} = \frac{7}{2} = 3.5$$

Ejercicio 3 *La vida en horas de un tipo de lámpara de radio es una variable aleatoria cuya función de densidad es*

$$f(x) = \begin{cases} 0 & si \quad x < 100 \\ \frac{k}{x^2} & si \quad x \geq 100 \end{cases}$$

(a) *(0.3 puntos) Determina el valor de k.*

Solución.

$$1 = \int_{100}^{+\infty} \frac{k}{x^2}\,dx = -k\cdot x^{-1}\Big|_{100}^{+\infty} = \frac{k}{100} \Rightarrow k = 100$$

(b) *(0.3 puntos) ¿Qué porcentaje de lámparas no superan las 500 horas de vida?*

Solución.

Sea X la variable aleatoria "tiempo de vida de una lámpara".

$$P(X \leq 500) = \int_{100}^{500} \frac{100}{x^2}\, dx = -100 \cdot x^{-1}\big|_{100}^{500} = \frac{4}{5} = 0.8$$

Un 80 % de las lámparas no superan las 500 horas de vida.

(c) *(0.3 puntos) Determina el número de horas de vida que no son superadas por exactamente el 20 % de las lámparas.*

Solución.

$$0.2 = P(X \leq q_{0.2}) = \int_{100}^{q_{0.2}} \frac{100}{x^2}\, dx = -100 \cdot x^{-1}\big|_{100}^{q_{0.2}} = 1 - \frac{100}{q_{0.2}} \Rightarrow q_{0.2} = 125$$

En un aparato de radio hay 5 de estas lámparas que funcionan de modo independiente.

(d) *(0.3 puntos) Calcula la probabilidad de que ninguna lámpara del aparato de radio tenga que ser sustituída durante las primeras 300 horas de uso.*

Solución.

Sea Y la variable aleatoria "número de lámparas que duran más de 300 horas".

$Y \sim Bin(n = 5, p)$

Siendo

$$p = P(X > 300) = \int_{300}^{+\infty} \frac{100}{x^2}\, dx = -100 \cdot x^{-1}\big|_{300}^{+\infty} = 1/3$$

$$P(Y = i) = \binom{5}{i} \left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{5-i}, \quad i = 0, 1, 2, 3, 4, 5.$$

$$P(Y = 5) = \binom{5}{5} \left(\frac{1}{3}\right)^5 = \frac{1}{3^5}$$

(e) *(0.3 puntos) ¿Cuál es el número medio de lámparas que deben cambiarse durante las primeras 300 horas de uso del aparato de radio?*

Solución.

Sea Z la variable aleatoria "número de lámparas que deben cambiarse durante las primeras 300 horas".

$Z \sim Bin(n = 5, p = 2/3)$

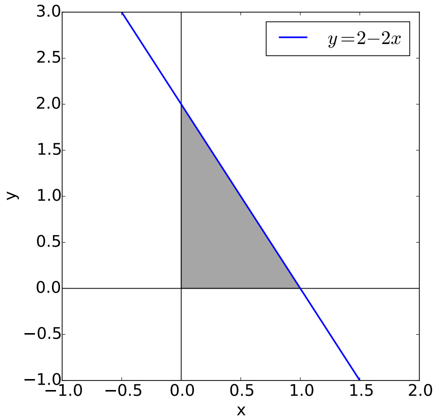
$$E(Z) = np = 10/3$$

Ejercicio 4 *La función de densidad de probabilidad conjunta de dos variables aleatorias X e Y es*

$$f_{(X,Y)}(x,y) = \begin{cases} C & \textit{si } \; 0 \le x \le 1, \;\; 0 \le y \le 2(1-x) \\ 0 & \textit{en el resto} \end{cases}$$

(a) *(0.2 puntos) Dibuja la región donde $f_{(X,Y)}$ no es cero.*

Solución.



(b) *(0.4 puntos) Calcula la constante C.*

Solución.

$$1 = \int_0^1 dx \int_0^{2-2x} C\, dy = C \int_0^1 (2-2x)\, dx = C$$

(c) *(0.4 puntos) Calcula las densidades marginales. ¿Son las variables aleatorias X e Y independientes?*

Solución.

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\, dy = \begin{cases} \int_0^{2-2x} dy = 2-2x & \text{si } x \in (0,1) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\, dx = \begin{cases} \int_0^{\frac{2-y}{2}} dx = \frac{2-y}{2} & \text{si } y \in (0,2) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_X(x) \cdot f_Y(y) = \begin{cases} \frac{(2-2x)(2-y)}{2} & \text{si } x \in (0,1), \;\; y \in (0,2) \\ 0 & \text{en otro caso} \end{cases}$$

Luego $f_X(x) \cdot f_Y(y) \neq f_{(X,Y)}(x,y)$, y X e Y no son independientes.

(d) *(0.6 puntos) Calcula $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ y $E[X \cdot Y]$.*

Solución.

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx = \int_0^1 (2-2x)x\, dx = \frac{1}{3}$$

$$E[Y] = \int_{-\infty}^{+\infty} y f_Y(y)\, dy = \int_0^2 \frac{2-y}{2} \cdot y\, dy = \frac{2}{3}$$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)\, dx = \int_0^1 (2-2x)x^2\, dx = \frac{1}{6}$$

$$E[Y^2] = \int_{-\infty}^{+\infty} y^2 f_Y(y)\, dy = \int_0^2 \frac{2-y}{2} y^2\, dy = \frac{2}{3}$$

$$\begin{aligned} E[X \cdot Y] &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} xy f_{(X,Y)}(x,y)\, dy = \int_0^1 dx \int_0^{2-2x} xy\, dy \\ &= \frac{1}{2}\int_0^1 x(2-2x)^2\, dx = \frac{1}{2}\int_0^1 (4x + 4x^3 - 8x^2)\, dx = \frac{1}{6} \end{aligned}$$

(e) *(0.4 puntos) Calcula la matriz de covariancias de (X, Y).*

Solución.

$$Var[X] = E[X^2] - (E[X])^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$$

$$Var[Y] = E[Y^2] - (E[Y])^2 = \frac{2}{3} - \frac{4}{9} = \frac{2}{9}$$

$$Cov[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y] = \frac{1}{6} - \frac{2}{9} = -\frac{1}{18}$$

La matriz de Covarianzas de (X, Y) será

$$\begin{bmatrix} \frac{1}{18} & -\frac{1}{18} \\ -\frac{1}{18} & \frac{2}{9} \end{bmatrix}$$

Ejercicio 5 *(1 punto) De un parámetro desconocido θ, del que depende la distribución de la variable aleatoria X, se sabe que, dada una muestra de tamaño n, el estadístico $\dfrac{\theta - \overline{X}}{2+n}$ se distribuye como una t de Student con $n+1$ grados de libertad. Calcula un intervalo de confianza para θ, con un nivel de confianza del 86 %, siendo 15, 17, 13, 21, 23, 14, 16 una muestra de X.*

Indicación: *Si Y es una variable aleatoria t de Student con 8 grados de libertad se cumple*

y	-1.638	-1.159	-0.182	1.159	1.638
$P(Y \leq y)$	0.07	0.14	0.43	0.86	0.93

Solución.

$n = 7 \Rightarrow \dfrac{\theta - \overline{X}}{2+7} \sim t_8.$

Por ser el nivel de confianza $0,86 = 1 - \alpha \Rightarrow \dfrac{\alpha}{2} = 0.07$. Los cuantiles 0.07 y 0.93 de un t de Student con 8 g. de l. son -1.638 y 1.638, respectivamente. Entonces,

$$0.86 = P\left(-1.638 \leq \frac{\theta - \overline{X}}{9} \leq 1.638\right) =$$

$$= P\left(\overline{X} - 1.638 \cdot 9 \leq \theta \leq \overline{X} + 1.638, \cdot 9\right)$$

$$\overline{x} = \frac{1}{7}(15 + 17 + 13 + 21 + 23 + 14 + 16) = 17$$

$$I.C._{0.86}(\theta) = 17 \pm 1.638 \cdot 9 = (2.258,\ 31.742)$$

Ejercicio 6 *(1 punto)*

En una entidad financiera la ganancia, en miles de euros, que se obtiene con una inversión a un año de un millón de euros sigue una distribución normal con una ganancia media de 50, pero con una desviación típica desconocida, σ. Se toma una muestra aleatoria de tamaño 10 y se obtiene que $s_1 = 5.62$. Contrasta, con un nivel de significación $\alpha = 0.08$, si $\sigma = 6$ frente a la hipótesis alternativa $\sigma \neq 6$ ¿Cuál es el p-valor del contraste? ¿Para qué niveles de significación podrías aceptar la hipótesis nula?

Indicación 1: *Si se verifica que $\sigma = \sigma_0$, el estadístico $\frac{(n-1)S_1^2}{\sigma_0^2}$ tiene una distribución χ^2_{n-1}.*

Indicación 2: *Si Y es una variable aleatoria χ^2 con 9 grados de libertad se cumple*

y	*3.105*	*3.866*	*7.628*	*7.896*	*8.541*	*17.608*
$P(Y \leq y)$	*0.04*	*0.08*	*0.428*	*0.455*	*0.519*	*0.96*

Solución.

El contraste es

$$H_0 : \sigma = 6$$
$$H_1 : \sigma \neq 6$$

Supuesto que es cierta la hipótesis nula el estadístico $\frac{(10-1)S_1^2}{6^2}$ se distribuye como una χ^2_9. Por ser $\alpha = 0.08$, los cuantiles $\frac{\alpha}{2} = 0.04$ y $1 - \frac{\alpha}{2} = 0.96$ de una v.a. χ^2_9 son 3.105 y 17.608, respectivamente.

En nuestro caso $\frac{(10-1)s_1^2}{6^2} = 7.896 \in (3.105,\ 17.608)$, por lo que no se rechaza la hipótesis nula.

$P(\chi^2_9 \leq 7.8961) = 0.455$ y $P(\chi^2_9 \geq 7.8961) = 0.545$, luego $\min\{0.455,\ 0.545\} = 0.455$.
Así, el p-valor es $2 \cdot 0.455 = 0.91$, por lo que se aceptaría la hipótesis nula para niveles de significación inferiores a 0.91.

Ejercicio 7 *Sea $W(n)$ ruido blanco, es decir $W(n)$, $n = 0, 1, \ldots$, son variables aleatorias independientes e identicamente distribuidas. Supóngase que las variable aleatorias $W(n)$, $n = 0, 1, \ldots$, tienen distribución $N(0, 1)$.*

(a) *(0.5 puntos) Calcula la autocorrelación $R_X(n, n+k)$ del proceso $X(n) = W(n+1) - W(n)$ para $k = 0, 1, \ldots$*

Solución.

Cuando $n = m$ se tiene que $E[W(n)W(m)] = E[W(n)^2] = \mathrm{VAR}[W(n)] = 1$. Cuando $n \neq m$ se tiene que $E[W(n)W(m)] = E[W(n)]E[W(m)] = 0$ ya que $W(n)$, $W(m)$ son independientes y tienen media

0. Entonces

$$
\begin{aligned}
&R_X(n,n+k) = E\big[X(n)X(n+k)\big] = E\Big[\big(W(n+1)-W(n)\big)\big(W(n+1+k)-W(n+k)\big)\Big] \\
&= E\big[W(n+1)W(n+1+k)\big] - E\big[\big(W(n+1)W(n+k)\big] \\
&\quad - E\big[\big(W(n)W(n+1+k)\big] + E\big[W(n)W(n+k)\big] \\
&= \begin{cases} 2, & k=0 \\ -1, & k=1 \\ 0, & k=2,3,\ldots \end{cases}
\end{aligned}
$$

(b) *(0.5 puntos) Determina las distribuciones de segundo orden del proceso $W(n)$*

Solución.

Si $n \neq m$ entonces $W(n)$ y $W(m)$ son independientes y son normales. Por tanto, si $n \neq m$, la distribución de $[W(n), W(m)]$ es normal bidimensional con vector de medias $(0,0)$ y matriz de covarianzas

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(c) *(0.5 puntos) Calcula $P\big[W(0)+W(1)+W(2) < 1\big]$*

Solución.

La variable aleatoria $W(0)+W(1)+W(2)$ tiene distribución normal de media 0. Utilizando que $W(0), W(1)$ y $W(2)$ son independientes se obtiene que

$$\mathrm{VAR}\left[W(0)+W(1)+W(2)\right] = \mathrm{VAR}\left[W(0)\right] + \mathrm{VAR}\left[W(1)\right] + \mathrm{VAR}\left[W(2)\right] = 3$$

Entonces

$$P\big[W(0)+W(1)+W(2) < 1\big] = P\left(\frac{W(0)+W(1)+W(2)}{\sqrt{3}} < \frac{1}{\sqrt{3}}\right) = \Phi\left(\frac{1}{\sqrt{3}}\right) \approx 0.7181$$